EXPLORING THE LATENT STRUCTURE OF COLLABORATIONS IN MUSIC RECORDINGS: A CASE STUDY IN JAZZ

Nazareno Andrade
Universidade Federal de Campina Grande

Flavio Figueiredo
IBM Research - Brazil

ABSTRACT

Music records are largely a byproduct of collaborative efforts. Understanding how musicians collaborate to create records provides a step to understand the social production of music. This work leverages recent methods from trajectory mining to investigate how musicians have collaborated over time to record albums. Our case study analyzes data from the Discogs.com database from the Jazz domain. Our analysis examines how to explore the latent structure of collaboration between leading artists or bands and instrumentists over time. Moreover, we leverage the latent structure of our dataset to perform large-scale quantitative analyses of typical collaboration dynamics in different artist communities.

1. INTRODUCTION

Collaboration is a major component of musical creation. Examining who has collaborated in a record is a common method to understand their style, content, and process of creation. Collaborators leave a mark in the music, and may affect the style of the leading artists themselves. For example, the fact that Miles Davis collaborated with Charlie Parker in the beginning of his career can be seen as an important influence in the development of his style.

Looking at a larger picture, understanding the string of collaborators of a musician over his or her career is also prolific source of information to understand the career itself. Reusing the same example, it is possible to partly describe changes in Miles Davis' style in the 70s by describing how he changed the musicians recording with him. At the same time, similarities in the sequence of collaborators for two artists may denote similarities in the artists themselves. Complementarily, identifying common sequences of leading artists with which different instrumentists have performed also helps understanding how styles and communities of musical creation evolve.

From a quantative standpoint, collaboration patterns have often been studied through the use of methods from graph analysis to large-scale collaboration networks (e.g. [1, 9, 16, 18, 21]). However, although these methods provide valuable insights, they fail to focus on longitudinal views of collaboration. This way, they do not allow for examining patterns in collaboration trajectories.

This work leverages recent methods proposed for mining trajectories in object consumption to study collaborations trajectories among musicians. We use TribeFlow [7], a method recently shown to accurately and expressively discover latent spaces of consumption sequences in the Web domain [7]. Our work explores how this model can also be used to discover latent structures in the trajectories of musicians as they collaborate with the leading artists or bands in records. This exploration is done through a case study with Jazz records. Collaborators and musicians as extracted from the Discogs.com collaborative database of discographic information.

The rest of this paper is organized as follows. In the next section we discuss related work. An overview of the TribeFlow model is described in Section 3. This is followed by a description of our datasets in Section 4. Our main results are discussed in Sections 5 to 7. Section 5 discusses the latent trajectories (collaboration spaces) extracted with TribeFlow. Here, we discuss how the method extracts a semantically meaningful latent representation of our datasets. In Section 6 we compare the collaboration spaces of different artists. Section 7 discusses how artists move between collaboration spaces over time. Finally, in Section 8 we conclude the paper.

2. RELATED WORK

Our cultural products, music being no exception, are strongly tied of our social interactions, and in particular to the dynamics of such interactions. Realizing the importance of understanding networks of interacting collaborators, various research efforts have looked into large-scale creation, dissemination and curation of information by groups of individuals [2,3,5,10–12,15,17,20,21]. Some efforts have also specifically focused on understanding musical recordings as a collaborative effort [1, 8, 9, 16, 18, 19]. Nevertheless, much less attention has been given to the dynamics of collaborations trajectories as we do.

With regards to musical production, very recently Bae et. al. [1] looked into the network properties and community structure of the ArkivMusic [1] database. This database, contains meta-data on classical music records. The authors looked into complex network properties such as power-law distributions and the small world effect [5] that exist

[1] `http://www.arkivmusic.com`

Attribution: Nazareno Andrade, Flavio Figueiredo. "Exploring The Latent Structure of Collaborations in Music Recordings: A Case Study in Jazz", 17th International Society for Music Information Retrieval Conference, 2016.

in such networks. Similar studies based on complex networks has also been performed for Brazilian music [9, 18], as well as contemporary popular music [16].

Regarding the community structure of collaboration networks, again the work of Bae et. al. [1] performed an analysis of the ego-network of contributors. Their analysis uncovered strength of social connections from performers of classical musical. In a similar note, Gleiser and Danon also looked into communities of jazz musicians [8].

Our approach in this paper complements the above by viewing collaborations as dynamic trajectories, not static networks as was done previously. This novel way of looking into collaboration employs recent advances in trajectory mining [7]. Viewing collaborations as trajectories can aid practitioners in understanding latent structures that reflect on the evolving career of a musician.

Finally, we point out that various previous efforts looked into the listening behavior of users as trajectories [7, 13, 14]. To perform our study, we employ the TribeFlow method [7]. This method has been shown to be more accurate and interpretable than state-of-the-art baselines in user trajectory mining [7]. We describe the method in the next section.

3. MINING COLLABORATION TRAJECTORIES WITH TRIBEFLOW

Music records (albums) represent a collaborative effort from different individuals such as band members, producers, hired instrumentists, and even graphical artists responsible for the artwork. Whenever individuals collaborate they leave a trail in their career. For example, in *2005 Lucas Dos Prazeres* collaborated with *Naná Vasconcelos* to create the album *Chegada*. Our goal in this study is to understand the collaboration trajectory of individuals. In this context, a trajectory is an ordered sequence of collaborations by a collaborator.

For the sake of clarity, we differentiate between the *artist* or *band* leading a record and the *collaborators* who participate in this album. These collaborators can themselves be the leading artists in other records. Paul, John, Ringo and George are all viewed as collaborators in an album by the artist The Beatles.

More formally, each trajectory defines the sequence of artists (ordered by time) that the collaborator contributed with. Let us define an album as a timestamp, artist/band, and a set of collaborators. That is, $r = (t_r, n_r, a_r, \mathcal{L}_r)$, where r is a record or album, t_r is a timestamp, n_r is the name of the album, a_r is the artist/band and $\mathcal{L}_r$ is the set of collaborators which contributed to r. The subscript r identifies the release for each element of the tuple. Let $\mathcal{R}$ be the set of records. Also, let us define that records are identified by integers $[1, |\mathcal{R}|]$, as well as that for any pair of records $t_i \leq t_{i+1}$. That is, records are ordered by the release timestamp and their ids correspond to the position of the record on the defined ordering.

With the definitions above, the trajectory of a collaborator c is defined as $T_c =< ..., a_i, a_{i+j}, ... >$, where for any pair a_i, a_{i+j} with $j \geq 1$, $t_i \leq t_{i+j}$ (by definition, albums ids are defined by their ordered timestamps). Also, $c \in L_i$ as well as $c \in L_{i+j}$. More importantly, we focus our study on the changes in collaborations over time. That is, we enforce $a_i \neq a_{i+j}$. With this choice, trajectories represent the changes in artists chosen by a collaborator over time.

To exemplify a trajectory, let's us look into John Coltrane as a collaborator. In 1955 to 1956, Coltrane collaborated on various recordings by Miles Davis. Later, in 1957 Coltrane collaborated with Thelonious Monk, again, in various recordings. Afterwards, John Coltrane returned to collaborate with Miles Davis in 1958. Taking this small slice of time as an example, the trajectory of John Coltrane would be represented as: $<$ Miles Davis, Thelonious Monk, Miles Davis $>$. Notice that, regardless of partaking in many records with Miles Davis in 1955 and 1956, the trajectory only captures the change in collaboration from Davis to Monk.

It is important to notice that there exists a variety of latent factors that lead to a collaboration. That is, while some collaborations will emerge due to geographical constraints, others may exist because of musical genres, temporal influence, or social network factors. The trajectory, as defined above, will essentially exist because of choices by the collaborator to collaborate with the artist motivated by these factors. Thus, the end result, regardless the of the underlying factors, is always the same a trajectory of trails/choices (artists) that as collaborator has worked with.

3.1 The TribeFlow Model

To extract the latent structure of trajectories, we employ TribeFlow [7], a recent method proposed to mine sequential data. TribeFlow has recently been shown to discover a meaningful latent structure in a variety of different settings. Here, we apply the method to understand musical collaborations based on the trajectories T_c.

In our setting, TribeFlow models collaborations as random choices over random environments by a collaborator. One example of a random environment can be: Jazz Artists from New Orleans in the 1960s. Due to various constraining factors, as explained above, a collaborator will choose to play with an artist from this environment over a set of albums. After recording these albums, the collaborator will again choose an environment (in some cases, the same as before) and move on to record more albums with different artists. Thus, trajectories are captured as random choices (or random walks) over random environments. Each environment captures a latent factor that leads to collaborations between collaborators and artists.

TribeFlow works using as input a set of trajectories. Given the set of collaborators $c \in \mathcal{C}$, the set of artists $a \in \mathcal{A}$, as well as the set of records $r \in \mathcal{R}$, TribeFlow will explore as input the total set of trajectories $T_c \in \mathcal{T}$. r was defined above. a and c can be defined as the names of artists and collaborators, respectively. A single parameter is required to execute the model. This parameter $k = |\mathcal{Z}|$ captures the number of latent environments $z \in \mathcal{Z}$.

TribeFlow defines a Bayesian graphical model (ommited due to space, see [7]) that learns by performing

Gibbs sampling for each entry (a_{i+j}) in every trajectory $\mathcal{T}$ from the following posterior:

$$P[z|c, a_{i+j}, a_i] = \frac{P[z|c]P[a_i|z]P[a_{i+j}|z]}{1 - P[a_i|z]} \quad (1)$$

$P[a|z]$, $P[z|c]$ and $P[z]$ are all probability distributions estimated with TribeFlow. After the model is trained, the above probabilities can be exploited to answer various queries, as we now describe.

3.2 Answering Questions with TribeFlow

First, we point out TribeFlow's model is highly interpretable, since it represent probabilities over $\mathcal{C}$, $\mathcal{A}$, and $\mathcal{Z}$.

By employing the graphical model described in [7], we can re-arrange the probability equations to answer various questions using the TribeFlow model. More specifically:

What is the probability that a collaborator goes from on to collaborate with artist a after choosing environment z? This initial likelihood is captured by the probability $P[a|z]$, learned by the model. It captures the importance of artists in a given environment, that is, artists with high $P[a|z]$ will likely attract more collaborators within that environment.

What is the probability that a collaborator goes from collaborating with artist a to artist a'? Given that a collaborator will collaborate with various artists over a trajectory, this first question captures the importance of artists as links to other artists. That is, how likely is a collaborator to follow-up on his/her career with a' after playing with a. This question is assessed by:

$$P[a'|a] = \sum_{z \in \mathcal{Z}} P[a'|z]P[a|z]P[z] \quad (2)$$

What is the probability of collaborating with environment z' after collaborating with environment z? This question is similar to the previous one. However, it captures the notion of transitions between environments. For instance, how likely is it that a collaborator will go from playing with Dixieland artists to playing with Bebop artists. $P[z'|z]$ is thus defined as:

$$P[z'|z] = \sum_{a \in \mathcal{A}} P[a|z']P[z']P[a|z] \quad (3)$$

What is the probability that an environment z caused collaborator to go from playing with a to playing with a'? This final question can be used to explain trajectories. It capture's how likely is an environment z to cause a change in collaboration from a to a'. This final question is answered with the posterior equation above (Eq. 1).

3.3 Learning the Model

Finally, we point out that our results were achieved by executing the method with the same parameters as discussed by the authors in [7]. More importantly, we made use of the TribeFlow **without** employing the inter-event time heuristics discussed in [7]. We found that employing such heuristics had little to no effect on our results. This effect most likely happens because timestamps t are usually expressed in years (e.g., 2005) on the Discogs dataset. For this reason, on the data we analyzed from 15% to 30% of collaborations happen within a single year, making the time between collaborations useless in such cases (they are all zero). Also, frequent collaborations can happen both over short and large periods of times, such as individuals that take hiatuses on their careers.

Given the exploratory nature of our work, for the sake of interpretability, all analysis in this work use $|\mathcal{Z}| = 30$. This number of latent environments provides an expressive range of latent factors for our purposes while keeping sensemaking easily tractable. To understand how to fine tune $|\mathcal{Z}|$ for other tasks (e.g., prediction) see [6, 7].

Table 1. Summary of the dataset used.

Releases	#		54,466
Artists	#		23,890
	in degree	median	5
		mean	14.8
		max	3,052
Collaborators	#		70,320
	out degree	median	1
		mean	5
		max	830
Collaborations	#		352,932

4. DATA USED

To investigate collaboration among musicians, we leverage the Discogs.com database. Discogs is a collaborative site to register and annotate discographies which makes its database freely available. At the time of writing, the database registers approximately six million record releases, including multiple releases of a same record (eg. CD and LP or LPs releases in different countries). Part of this data is annotated with genre and more specific style tags, and with information about which collaborators participated in a record and in which capability. For example, it is registered in the database that Ron Carter played the bass in Charles Mingus's record Three of Four Shades of Blues, initially released in 1977. As this data is collaboratively created, it is likely biased towards the interests of its contributors, and it is naturally incomplete. Nevertheless, due to its sheer volume and to the community verification of its information, this database provides a promising data source for investigating how collaboration patterns can be understood through trajectory mining.

For this purpose, we use a dataset extracted from the Discogs database as a case study in collaboration. The dataset is comprised of all records tagged with the Jazz genre and which possess metadata about instrumentists in

Table 2. Six exemplary latent environments found learned through TribeFlow from the data. For each environment, we present the first collaborators and artist most strongly associated with the environment. Collaborators are listed first; artists are in italics. All labels were given by the authors based on the artists and collaborators listed.

(a) Bebop	***(b) Bebop 2***	***(c) Free Jazz***
Dizzy Gillespie, Miles Davis, J.J. Johnson, Charlie Rouse, Lucky Thompson, Charlie Parker *Dizzy Gillespie, Charlie Parker, Miles Davis, Thelonious Monk, Dizzy Gillespie And His Orchestra*	John Coltrane, Freddie Hubbard, Donald Byrd, Hank Mobley, Lee Morgan *Art Blakey & The Jazz M., Miles Davis, John Coltrane, Art Blakey, Charles Mingus, Max Roach*	Steve Lacy, Don Cherry, Archie Shepp, Roswell Rudd Lester Bowie *Archie Shepp, The Sun Ra Arkestra, Cecil Taylor, Steve Lacy Anthony Braxton, Ornette Coleman*
(d) Improvisation UK Big Bands	***(e) Italian***	***(f) Fusion***
Paul Rutherford, Evan Parker, John Edwards, Johannes Bauer, Paul Rogers, Malcolm Griffiths *London Improvisers Orchestra, Chris McGregor's Brotherhood Of Breath, Globe Unity Orchestra*	Giovanni Maier, Gianni Basso, Lauro Rossi, Dino Piana, Giancarlo Schiaffini *Giorgio Gaslini, Italian Instabile Orchestra, Enrico Rava, Nexus, Chet Baker Nicola Conte*	Don Alias, David Liebman, Dave Holland, Joe Lovano, Bob Berg, Mino Cinelu *Miles Davis, Jaco Pastorius, John Scofield, Chick Corea, Mike Stern, Herbie Hancock*

the Oct 2015 database dump. After removing duplicated releases and releases with no collaboration metadata, the release name, year, and collaboration data were extracted from each release. Furthermore, we focus on instrumentists in which the metadata associated with his/her role in the record contained one of the following words: *bass, guitar, drum, vocal, voic, percuss, keyboard, trumpet, sax, saxophon, trombon, flute, synthes*.

The data resulting from this process is summarized in Table 1. Considering a collaboration as a *(collaborator, time, leading artist, record)* tuple, the in degree of an artist a denotes the number of distinct tuples where a is present in the data. Similarly, the out degree of a collaborators c denotes the number of distinct tuples in which c is present. Ron Carter and Miles Davis are the collaborators and artist with the largest degree in the data. Inspecting the releases, it is possible to note that the numbers of popular artists are slightly inflated due to compilation releases.

5. LATENT TRAJECTORY SPACES

Each latent environment found by TribeFlow can be seen as the result of a set of latent factors that influence the movement of collaborators between artists. Inspecting the collaborators and artists most associated with each environment thus sheds light on what are the relevant factors affecting collaboration sequences in our dataset.

The latent environments found in our case study reveal clear loadings on the environments of stylistic, geographical and chronological latent constraints that shape collaboration trajectories. Table 2 shows six exemplary latent environments as described by the collaborators and artists (in italics) most associated with them [2].

It is possible to clearly distinguish in the first four environments the styles of jazz most often associated with the artists, and to note that in several cases collaborators have notoriously recorded with multiple artists in that environment. It is worthwhile noting that we remove collaborations where the collaborator and artist are equal. For example, there are recordings in the data both of Dizzy Gillespie collaborating in Miles Davis albums and vice-versa.

As for the chronology, it is possible to see in the examples that collaborations of a same period are loaded in different environments. For example, environments *(a)*, *b* and *f* are all associated with Miles Davis. However, his collaborators are divided in these three spaces according to the period of the collaboration. Most markedly, it is possible to distinguish his collaborators from the 70s in environment *(f)* versus earlier collaborators in environment *(b)* and even earlier on *(a)*.

Similarly, the set of top collaborators listed in the *(d)* environment is a core part of the three bands listed as artists. Moreover, in this case, as in the environment *e*, there is a relation of collaborators, artists and geography: the three artist groups listed in environment *(d)* are largely based in the UK, while artists in environment *(e)* are mostly Italians. Along the same lines, there are latent environments not listed in Table 2 that group Brazilian or Scandinavian jazzists, among others.

Before continuing, we point out that various efforts looked into the trajectories of users listening to music [7, 13, 14]. In our study, we make use of TribeFlow [7], a recent trajectory mining technique that has been shown to be both accurate and interpretable when compared to other state-of-the-art methods. We describe the method in the next section.

Besides looking at the intuitive similarities of collaborators or artists associated with an environment, a second possibility for sensemaking is to look for less obvious associations. For example, the presence of Chet Baker, an artist mostly known for his work in the USA, in envi-

[2] Access to the list environments and probabilities is available online at: `https://github.com/flaviovdf/tribeflow/` on the folder `scripts/ismir2016annotated`.

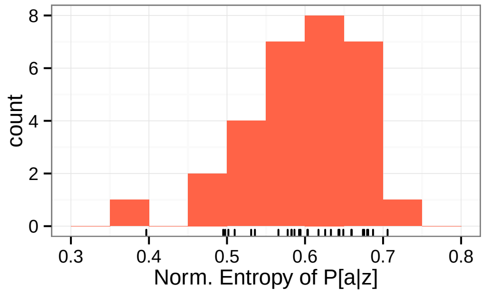

Figure 1. Normalized entropy of $P[a|z]$. The normalized entropy was computed for every z

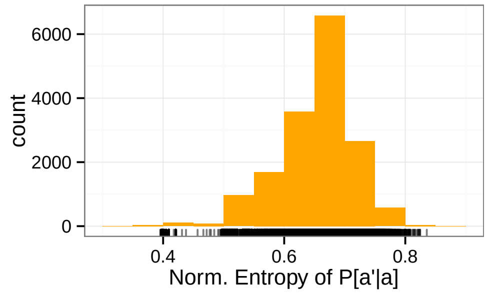

Figure 2. Normalized entropy of $P[a'|a]$. The normalized entropy was computed for every $a' \neq a$.

ronment *(e)* is an example of a relationship less obvious for some. An investigation of Chet Baker's collaborators shows that during his work and late life in Europe, Chet Baker recorded more than one album with a cast of mostly Italian musicians. Several of these musicians (eg. Enrico Rava) have trajectories of collaboration that touch on other Italian artists in Table 2.

6. DIFFERENCES IN COLLABORATION SPACES

Because TribeFlow has a probabilistic interpretation, it allows an analyst to investigate differences in probabilities learned for transitioning to/from different artists and latent environments. A relevant tool for doing so in our context is to examine the entropy of the probability distributions extracted with TribeFlow.

Along those lines, we first investigate which latent environments have a high/low entropy [4] in $P[a|z]$. Recall that, entropy captures the expected uncertainty in a probability distribution. Higher values of entropy indicate that a discrete distribution, our case, is closer to being uniform. Lower values of entropy indicate that the distribution is skewed to a subset of artist in our case.

In other words, the entropy of $P[a|z]$ captures the notion that after choosing to collaborate an artist from z, what is the uncertainty of choosing an artist. Environments with higher entropy indicate that most collaborations remain within a small subset of the artists.

In the following, we use the normalized entropy. Normalization is performed dividing the entropy $\sum_a -P[a|z]log(P[a|z])$ for each space z in the model by that of a Uniform distribution over the same artists. Figure 1 displays the distribution of the values of the normalized entropies for each latent environment.

In our data, the latent environments associated with the highest values of entropy display a normalized entropy of approximately 0.7. These environments seem to be either associated with free or experimental jazz, or to be mostly formed by collaborators and artists from a specific region outside the USA. For example, environments *(d)* and *(e)* are among the highest entropy environments, together with an environment associated with North-European jazz and another one associated with experimental jazz fused with World Music. As for the environments with least entropy,

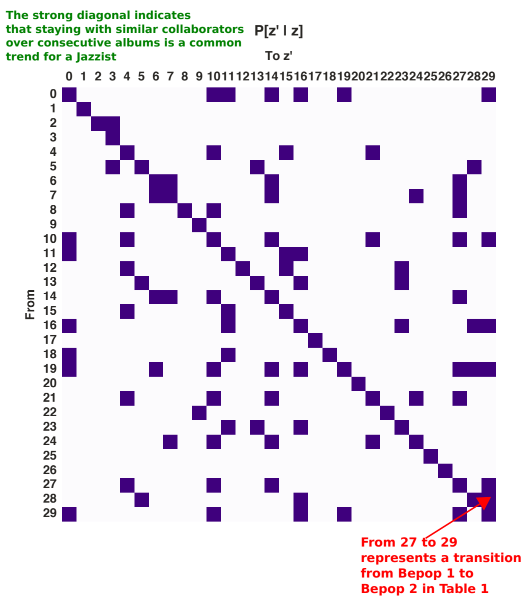

Figure 3. The $P[z'|z]$ highlighting transitions that are above uniform chance $(1/\mathcal{Z})$

the apparent pattern for top-3 environments is the combination of artists from older periods with the presence of a seminal figure. Namely, these three environments have Duke Ellington, Count Basie and Louis Armstrong as their top artists. This suggests that collaborators in our data associated with these latent environments had their trajectories gravitating around these artists with not much collaboration with the other artists in the same environments.

A similar approach can be used to identify who are the artists which have the highest entropy considering the probabilities of collaborating with other artists afterwards ($P[a'|a]$ - Eq. (2)). The five highest-entropy artists in this view are all jazzists often associated with avant-garde or free jazz: Anthony Braxton, Peter Brötzmann, Franz Koglmann, Herb Robertson, and Gerry Hemingway. In a sense, our model points that collaborators who record with these artists are follow no clear pattern in the following collaborations. This can be seen as a sign of the openess in the choice of these artists in collaborating. On the other end of the spectrum, musicians collaborating with artists in the former Czech Republic (eg. Czechoslovak Radio Jazz

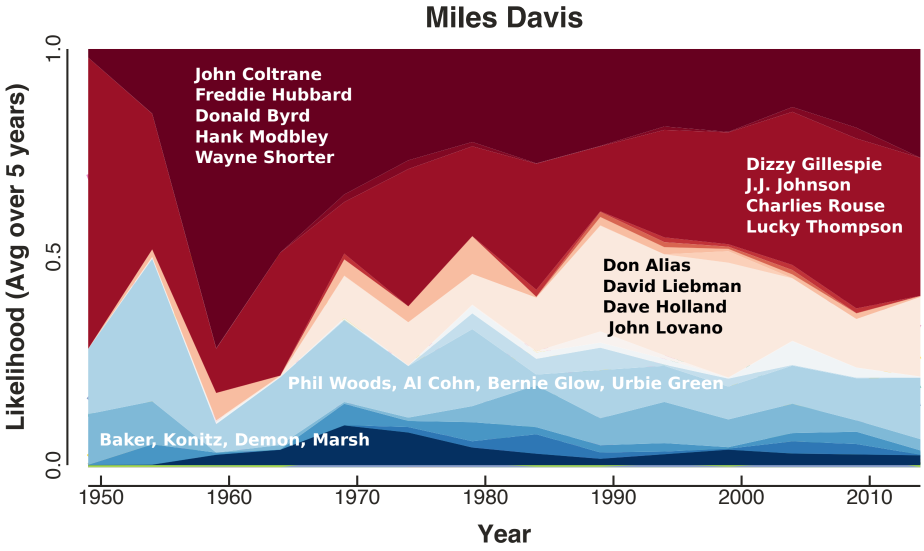

Figure 4. The career of Miles Davis as seen by the model.

Orchestra, Prague Big Band, Ji Vlek) are the ones displaying the least entropy. This last example is likely caused by geographical constraints.

7. MOVEMENT ALONG LATENT SPACES

Considering the probabilities of a collaborator transitioning between distinct spaces $P[z'|z]$ (Eq. (3)) gives us a view of which latent environments are likely to be adjacent in collaborators' trajectories. In Figure 3, we depict this transition matrix. For the sake of clarity, we highlight in this matrix only transitions that were above uniform random chance ($\frac{1}{Z}$). In this plot, latent spaces are numbered from 0 to $k-1$.

Observing the model fit to the data, we see that the pairs of latent environments with the highest probabilities of transition between them will be on the diagonals. That is, artists will likely remain collaborating within the same style after the previous records. Although this is expected, there interesting examples in the non-diagonals as well. For example, the highest probability in a non-diagonal happens between an environment where the most likely artist to collaborate with is Stan Kenton and a second environment where the correspondent artist is Woody Herman. These are two Big Band leaders who led popular bands during the first half of the 20th century. The following five highest probabilities follow a similar pattern, with the fifth being between environments *(b)* and *(a)* in Table 2.

A final frame in which we explore how to use TribeFlow in the context of collaboration trajectories is to inspect the trajectories associated with a prominent artist. Figure 4 shows the likelihood of an environment ($P[z|c, a_{i+j}, a]$ - Eq. (1)) of all collaborator transitions to reach Miles Davis over a period to be associated with each latent environment. We averaged this likelihood over every five years.

For this case, there is a marked change in the likely source of collaborators from the environments *(a)* to *(b)* from Table 2. This change correlates with a major change in the Miles Davis Quintet to the group that would compose it during the first half of the 60s. Wayne Shorter is one of the collaborators strongly associated with environment *(b)* who also recorded with multiple other artists associated with this environment, such as Freddie Hubbard.

8. DISCUSSION

Examining collaboration patterns is an important endeavor in understanding artists' influences and creations. Through a case study of collaboration in Jazz records, we have explored how to use TribeFlow to unveil latent structures in the trajectories of collaborators. The latent environments found in this case study are expressive and were able to help sensemaking in both popular and more niche collaboration groups in the data. Moreover, these environments seem to express at least stylish, chronological and geographical factors shaping collaboration trajectories. Due to the Bayesian approach of TribeFlow, it is possible to employ a direct probabilistic approach to investigate questions of association at different levels of relationship, such as artist to artist, and artist to environment.

Future work may extend the approach of this paper in deepening the link of the analysis conducted here in its musicological and historical aspects, employing a similar approach to other datasets, and extending both our modeling approach and the tools used so far to compare collaboration in different communities. Moreover, understanding how collaboration trajectories related to musical features (e.g., beat or tempo) can also help researchers better understand collaboration in recordings.

Acknowledgments: Research supported grant 460154/2014-1 from Brazil's National Research Council (CNPq).

9. REFERENCES

[1] Arram Bae, Doheum Park, Yong-Yeol Ahn, and Juyong Park. The multi-scale network landscape of collaboration. *PloS one*, 11(3):e0151784, 2016.

[2] Albert-László Barabási, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3):590–614, 2002.

[3] Alex Beutel, B Aditya Prakash, Roni Rosenfeld, and Christos Faloutsos. Interacting viruses in networks: can both survive? In *Proc. KDD*. ACM, 2012.

[4] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[5] David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[6] Flavio Figueiredo, Bruno Ribeiro, Jussara M Almeida, Nazareno Andrade, and Christos Faloutsos. Mining Online Music Listening Trajectories. In *Proc. ISMIR*, 2016.

[7] Flavio Figueiredo, Bruno Ribeiro, Jussara M Almeida, and Christos Faloutsos. TribeFlow: Mining & Predicting User Trajectories. In *Proc. WWW*, 2016.

[8] Pablo M Gleiser and Leon Danon. Community structure in jazz. *Advances in complex systems*, 6(04):565–573, 2003.

[9] Charith Gunaratna, Evan Stoner, and Ronaldo Menezes. Using network sciences to rank musicians and composers in brazilian popular music. In *Proc. ISMIR*, 2011.

[10] Paul Heymann and Hector Garcia-Molina. Collaborative creation of communal hierarchical taxonomies in social tagging systems. 2006.

[11] Brian Karrer and Mark EJ Newman. Competing epidemics on complex networks. *Physical Review E*, 84(3):036106, 2011.

[12] Paul Lamere. Social tagging and music information retrieval. *Journal of new music research*, 37(2):101–114, 2008.

[13] Andryw Marques, Nazareno Andrade, and Leandro Balby Marinho. Exploring the Relation Between Novelty Aspects and Preferences in Music Listening. In *Proc. ISMIR*, 2013.

[14] Joshua L. Moore, Shuo Chen, Thorsten Joachims, and Douglas Turnbull. Taste Over Time: The Temporal Dynamics of User Preferences. In *Proc. ISMIR*, 2013.

[15] Mark EJ Newman. Scientific collaboration networks. *Physical review E*, 64(1):016131, 2001.

[16] Juyong Park, Oscar Celma, Markus Koppenberger, Pedro Cano, and Javier M Buldú. The social network of contemporary popular musicians. *International Journal of Bifurcation and Chaos*, 17(07):2281–2288, 2007.

[17] Maximilian Schich, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing. A network framework of cultural history. *Science*, 345(6196):558–562, 2014.

[18] D de Lima Silva, M Medeiros Soares, MVC Henriques, MT Schivani Alves, SG de Aguiar, TP de Carvalho, G Corso, and LS Lucena. The complex network of the Brazilian Popular Music. *Physica A: Statistical Mechanics and its Applications*, 332, 2004.

[19] T Teitelbaum, P Balenzuela, P Cano, and Javier M Buldú. Community structures and role detection in music networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(4):043105, 2008.

[20] Jennifer Trant. Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1), 2009.

[21] Brian Uzzi and Jarrett Spiro. Collaboration and creativity: The small world problem. *American journal of sociology*, 111(2):447–504, 2005.